Matthews JNS, Forbes AB.

Stepped Wedge Designs: insights from a design of experiments perspective.

***Statistics in Medicine* (2017)**

DOI: https://doi.org/10.1002/sim.7403

Research Article

Statistics
in Medicine

Received XXXX

(www.interscience.wiley.com) DOI: 10.1002/sim.0000

Stepped Wedge Designs: insights from a design of experiments perspective

J. N. S. Matthews[a]*, A. B. Forbes[b]

Stepped wedge designs (SWDs) have received considerable attention recently, as they are potentially a useful way to assess new treatments in areas such as health services implementation. Because allocation is usually by cluster, SWDs are often viewed as a form of cluster-randomized trial. However, since the treatment within a cluster changes during the course of the study they can also be viewed as a form of crossover design. This article explores SWDs from the perspective of crossover trials, and designed experiments more generally. We show that the treatment effect estimator in a linear mixed effects model can be decomposed into a weighted mean of the estimators obtained from i) regarding a SWD as a conventional row-column design, and ii) a so-called vertical analysis, which is a row-column design with row effects omitted. This provides a precise representation of "horizontal" and "vertical" comparisons, respectively, which to date have appeared without formal description in the literature. This decomposition displays a sometimes surprising way the analysis corrects for the partial confounding between time and treatment effects. The approach also permits the quantification of the loss of efficiency caused by mis-specifying the correlation parameter in the mixed-effects model. Optimal extensions of the vertical analysis are obtained and these are shown to be highly inefficient for values of the within-cluster dependence which are likely to be encountered in practice. Some recently described extensions to the classic SWD incorporating multiple treatments are also compared using the experimental design framework.

Keywords: cluster randomized clinical trial; crossover design; linear mixed models; stepped wedge design

1. Introduction

Stepped wedge designs (SWDs) have recently received an increasing amount of attention, largely as a design for the evaluation of service delivery and implementation [1, 2]. The typical design, shown in Table 1 for six time periods and five sequences, denoted by SW(5), reveals the key features of the basic SWD: the trial is divided into successive time periods and into sequences, with the experimental units being randomly allocated to one of the sequences. All sequences start with the standard treatment in the first period and then change to the new treatment in a subsequent period, the change-over occurring at different times in the different sequences, but with the new treatment administered in the final

[a] *School of Mathematics & Statistics, Newcastle University, Newcastle upon Tyne, NE1 7RU, U.K.*
[b] *School of Public Health and Preventive Medicine, Monash University, Melbourne, VIC 3004, Australia*
* *Correspondence to: J.N.S. Matthews, School of Mathematics & Statistics, Newcastle University, Newcastle upon Tyne, NE1 7RU, U.K. Email: john.matthews@ncl.ac.uk*

Prepared using simauth.cls *[Version: 2010/03/10 v3.00]*

period of all sequences. The number of sequences in these designs corresponds to the number of 'steps', i.e. occasions on which the standard treatment changes to the new [3]. Features which make the design attractive for the assessment of service implementation are i) that all sequences start with the standard and finish with the new treatment and ii) once the changeover to the new treatment has occurred it is impossible to revert to the standard treatment.

	Time					
Sequence	1	2	3	4	5	6
1	0	1	1	1	1	1
2	0	0	1	1	1	1
3	0	0	0	1	1	1
4	0	0	0	0	1	1
5	0	0	0	0	0	1

Table 1. Standard stepped wedge design, SW(5), with $T = 6$ periods and $S = 5$ sequences comparing treatments 0 (standard) and 1 (new).

The experimental units are usually groups or clusters of patients, such as general practices, hospitals, care homes, rural communities or schools, rather than individual patients [1, 2]. Consequently SWDs are often viewed as a variant of the standard cluster-randomized trial. While many aspects of SWDs can profitably be seen in this light, the fact that they are divided into successive treatment periods and units are allocated to sequences of treatments, means that they also have much in common with crossover (CO) trials [4]. The purpose of this paper is to explore the extent to which the perspective of COs and, more generally, classical designed experiments, can illuminate statistical aspects of SWDs. For example, the literature on SWDs has loosely described 'horizontal' and 'vertical' comparisons (see, e.g., [5]) but without those being defined formally. We show that these can be defined explicitly using experimental design concepts and language.

In the next section we provide background on classical row-column designs and their relationship to SWDs. Section 3 provides a decomposition of the stepped wedge (SW) treatment effect under a linear mixed model with a variety of dispersion structures. Section 4 derives the variance of the SW treatment effect estimators under the correct and mis-specified models of the dispersion structure, and examines the impact of mis-specification. In Section 5 we derive optimal weights for the so-called 'vertical' analyses and Section 6 considers recently proposed extensions to multiple treatments and derives the relative efficiency of treatment effect estimators under different designs. The paper concludes with a discussion of the issues raised.

2. Stepped wedge designs as a classical row-column designs

The principal concerns for analyses of SWDs is that account should be taken of i) the fact that the designs extend over time and so there may be systematic time effects in the observed responses, and ii) that individual responses clustered within higher levels units are correlated. In the analysis of a CO study there are analogous concerns which are addressed by the use of period and patient effects. In addition the literature on CO designs often considers a term for the carryover effect of treatments but, for many medical applications, the appropriateness of the standard carryover model has been questioned (see [6]). For SWDs there are issues related to carryover [7] but they are of a somewhat different character and we will defer discussion of these until Section 7. Recent usage of CO designs in medical applications excludes carryover by using washout periods and in these circumstances a carryover term is not present in the model, so crossover designs are then essentially row-column designs, which we now describe.

In a row-column design, such as a Latin square or a Youden design, the aim is to arrange the treatments amongst the rows and columns so that i) the study can provide estimates of the treatment effects, notwithstanding the presence of row and column effects, and ii) that these estimates should have the best precision possible. Row-column designs in which

each treatment is replicated as nearly equally as possible within each row, and also within each column, will provide high precision estimates, and are often optimal with respect to appealing criteria. The most familiar example of an optimal row-column design is the Latin Square, but there are many more flexible row-column designs with optimal properties, such as Generalized Youden Designs (GYDs) (see, e.g. chapter 4 of [8]). Consider, for example, the two-treatment GYD shown in Table 2, where each treatment appears three times in each row and two or three times in each column (i.e. as nearly equally as possible): denote this design as NrO. Row and column effects can be eliminated from the treatment comparison in both SW(5) and NrO, but the more balanced layout of NrO allows this to be done more efficiently, and it is worth considering for a moment the details of how this is achieved.

	Time					
Sequence	1	2	3	4	5	6
1	1	0	1	0	1	0
2	0	1	0	1	0	1
3	1	0	1	0	1	0
4	0	1	0	1	0	1
5	1	0	1	0	1	0

Table 2. A Generalized Youden Design, denoted NrO, with $S = 5$ rows and $T = 6$ columns, having the same treatment replication as SW(5).

When contrasting SWDs and row-column designs a notation that is more commonly seen with the former than the latter will be adopted. Therefore the rows will be indexed by $s = 1, \ldots, S$, as these correspond with sequences in the SWD, and columns are indexed by $t = 1, \ldots, T$ as these correspond to successive periods, or time. If a single outcome Y_{st} were observed in row s and column t of the designs in Tables 1 and 2, then the usual model for a row-column design, adapted for the two-treatment case, is

$$Y_{st} = \alpha_s + \beta_t + D_{st}\theta + \eta_{st} \tag{1}$$

with fixed row effects α_s, fixed column effects β_t, treatment indicator $D_{st} \in \{0, 1\}$, and mutually independent errors η_{st} with mean 0 and variance σ^2_η. In the statistics literature, as well as the econometrics literature for panel data, such a model is called a two-way fixed effects model. The least squares estimator of the treatment effect θ is a linear combination of the observations,

$$\hat{\theta} = \sum_{s,t} a^{RC}_{st} Y_{st},$$

where the set of values a^{RC}_{st}, which we refer to as the coefficient matrix, are shown in Table 3 for the SW(5) and NrO designs. The superscript RC denotes that the coefficient matrix arises from the model (1), which includes row and column effects: the derivation of the a^{RC}_{st} is outlined in Appendix 1.

	Time					
Sequence	1	2	3	4	5	6
1	-10	14	8	2	-4	-10
2	-5	-11	13	7	1	-5
3	0	-6	-12	12	6	0
4	5	-1	-7	-13	11	5
5	10	4	-2	-8	-14	10

(a) Scaled coefficients for SW(5): above are $70a^{RC}_{st}$

	Time					
Sequence	1	2	3	4	5	6
1	2	-2	2	-2	2	-2
2	-3	3	-3	3	-3	3
3	2	-2	2	-2	2	-2
4	-3	3	-3	3	-3	3
5	2	-2	2	-2	2	-2

(b) Scaled coefficients for NrO: above are $36a^{RC}_{st}$

Table 3. Scaled coefficient matrices for designs SW(5) and NrO

The coefficients in Table 3 sum to zero down the columns, thereby eliminating column effects β_t, and the rows sum to

zero, eliminating the fixed row effects α_s from $\hat{\theta}$. Moreover, there is a symmetry to a standard stepped wedge design: if time were reversed in Table 1, then the design obtained would be the same as SW(5), but with 0s and 1s interchanged and the sequences written in the order 5 to 1. As the model (1) is temporally reversible, there is a corresponding anti-symmetry of the a_{st}^{RC} in Table 3(a), where reversing the order of the columns and of the rows yields the original matrix multiplied by -1. Inspection of the elements in Table 3(a) shows that the layout of SW(5) requires that the observations are combined with very different weights and in a rather unintuitive way: for example Y_{12} is given fourteen times the weight that is given to Y_{25}, and Y_{30} and Y_{36} are omitted altogether. This feature is attributable to the unbalanced nature of the SWD, in that treatment and time are partly confounded - something which is less marked for the more balanced design NrO, where the weights vary much less.

Also,

$$\begin{aligned}\mathsf{var}(\hat{\theta}) =& \sigma_\eta^2 \sum (a_{ij}^{RC})^2 \\ =& \tfrac{3}{7}\sigma_\eta^2 = 0.43\sigma_\eta^2 \quad \text{for SW(5)} \\ =& \tfrac{5}{36}\sigma_\eta^2 = 0.14\sigma_\eta^2 \quad \text{for NrO}\end{aligned}$$

Compared with NrO, which is the optimal design [8] for this number of rows, columns and treatments, the SWD has efficiency of only 32% (0.14/0.43), which quantifies the substantial loss of efficiency arising from the temporally imbalanced nature of the SWD.

If the model (1) is amended so that the row term α_s is omitted, then the coefficient matrix changes to a_{st}^C, as shown in Table 4 (the corresponding table for NrO is unchanged, as the treatment and row effects are orthogonal in this design). The variance of this estimator of θ is $\frac{1}{4}\sigma_\eta^2$, which is substantially smaller than $\frac{3}{7}\sigma_\eta^2$ (provided that σ_η^2 has the same value in both models, as would occur if the model including the α_s were inappropriate) and illustrates the penalty which has to be paid to eliminate simultaneously row and column effects in such an unbalanced design. This is despite the fact that the coefficients in Table 4 show that the first and last observations in each row are ignored in this estimator. Nevertheless, the unequal incidence of treatments in each column still exacts a price because this variance is nearly twice that for NrO.

	Time					
Sequence	1	2	3	4	5	6
1	0	4	3	2	1	0
2	0	-1	3	2	1	0
3	0	-1	-2	2	1	0
4	0	-1	-2	-3	1	0
5	0	-1	-2	-3	-4	0

Table 4. Scaled coefficient matrices for designs SW(5) when the model omits row effects: above are $20a_{st}^C$

The analysis implied by the coefficients in Table 4 compares the mean of the observations on treatment 0 with the mean of those on treatment 1, separately within each column. Under the amended model (1) the resulting T column-wise estimators of θ are independent and Table 4 essentially prescribes how these are to be combined, namely by being weighted inversely to their variances. Such an analysis closely resembles the the so-called *vertical* analyses mentioned in the context of SWDs, see [5]. As is remarked in [5], if observations in successive columns are dependent, it may be better if the separate column-wise contrasts are combined differently from the way implied in Table 4, a point taken up in Section 5.

Vertical analyses, being based on within-column contrasts, are sometimes recommended since they *preserve the randomization*, because in a SWD units are randomized to rows not columns. As mentioned in [5] this is also true of analyses based on the coefficients in Table 3(a), which can be viewed as aggregations of within-column contrasts, albeit

with less obvious weights and the rather counter-intuitive property that some observations receiving the same treatment receive weights of opposite sign.

3. The Stepped Wedge treatment effect estimator in a linear mixed model

3.1. The repeated cross-sectional sampling case

The link between the above discussion and SWDs as they are usually implemented is not immediate, because the formulation implied by (1) is not used for SWDs. In particular, clusters, not individuals, are allocated to sequences and responses within clusters are generally correlated. A more realistic formulation denotes the continuous response of an individual j, in the ith cluster allocated to sequence s, observed in the tth period by Y_{sijt}: the treatment administered in that period will depend on the sequence s to which cluster i has been allocated. While the same cluster is repeatedly measured throughout the study, the question arises of whether the individuals are measured on several occasions or just once. This will depend on context, as noted by Turner and colleagues [9] in the context of a cluster CO design. For the moment we assume that the application is such that each individual is measured just once, in a particular cluster and period, and as such there are separate samples of individuals in each period in each cluster. Also for simplicity it will be assumed that the same number of individuals, n, are observed in each period in each cluster, and that an equal number of clusters, N, is allocated to each of the S sequences in the design. A widely used model for the outcomes Y_{sijt} was proposed by Hussey and Hughes [10], namely

$$Y_{sijt} = \beta_t + D_{st}\theta + \alpha_{si} + \epsilon_{sijt}, \quad s = 1, \ldots, S;\ i = 1, \ldots, N\ j = 1, \ldots, n;\ , t = 1, \ldots, T \tag{2}$$

where the number of periods, T, will be $S + 1$ for designs such as those in Table 1, although this need not hold for variations on this design. In this model β_t is a fixed period effect, equivalent to the column effect, and D_{st} is as in (1), and θ is the treatment effect. The cluster effect is described by a random term α_{si} (as opposed to the fixed α_s in (1)), independent of the individual residuals ϵ_{sijt}; both terms have zero mean and variances τ^2 and ν^2, respectively.

Estimation of the β_t and θ can proceed directly using maximum likelihood if we assume that the random terms in (2) are Normally distributed, or otherwise using a generalized estimating equation (GEE). However, an equivalent and more helpful form of the equations for the estimation of β_t and θ can be derived by starting from the model implied for the cluster/period means, $\bar{Y}_{sit} = \sum_{j=1}^{n} Y_{sijt}/n$, namely

$$\bar{Y}_{sit} = \beta_t + D_{st}\theta + \alpha_{si} + \bar{\epsilon}_{sit} \tag{3}$$

from which $\mathsf{var}(\bar{Y}_{sit}) = \tau^2 + n^{-1}\nu^2$ and $\mathsf{cov}(\bar{Y}_{sit}, \bar{Y}_{sit'}) = \tau^2$, $t \neq t'$. Suppose $\bar{Y}_{si} = (\bar{Y}_{si1}, \ldots, \bar{Y}_{siT})^T$ denotes the T-dimensional vector of the means $\bar{Y}_{sit}$, $t = 1, \ldots, T$ and $\gamma^T = (\beta^T, \theta)$. Then if we write $\mathsf{E}(\bar{Y}_{sit})$ from (3) as $X_s\gamma$, where X_s is the design matrix for the fixed effects, which has a common value for all clusters i allocated to sequence s, then standard results on mixed models show that

$$\begin{aligned}\hat{\gamma} &= \left[\sum_{s=1}^{S}\sum_{i=1}^{N} X_s^T V^{-1} X_s\right]^{-1} \sum_{s=1}^{S} X_s^T V^{-1} \sum_{i=1}^{N} \bar{Y}_{si} \\ &= \left[\sum_{s=1}^{S} X_s^T V^{-1} X_s\right]^{-1} \sum_{s=1}^{S} X_s^T V^{-1} \bar{Y}_s,\end{aligned} \tag{4}$$

where $\bar{Y}_s = N^{-1}\sum_{i=1}^{N} \bar{Y}_{si}$. Here V is the variance of $\bar{Y}_{si}$, and although a general form could be envisaged, we focus on that induced by (3) which is $n^{-1}\nu^2 I_T + \tau^2 J_{T,T}$, where I_T is the $T \times T$ identity matrix and $J_{T,T'}$ is the $T \times T'$

matrix of ones. In the following it will often be convenient to write V as $\sigma^2((1-\psi)I_T + \psi J_{T,T})$, with $\sigma^2 = n^{-1}\nu^2 + \tau^2$ and $\psi = n\tau^2/(\nu^2 + n\tau^2)$: note that ψ, the correlation of the cluster-period means, depends on n. This is in contrast to the usual measure of dependence encountered in cluster randomized trials, the intra-cluster correlation (ICC), namely $\rho = \tau^2/(\nu^2 + \tau^2)$, the correlation between individuals in the same cluster that is induced by (2). Notice that $\psi = n\rho/[(1-\rho) + n\rho]$, so the dependence on n means that ψ can take substantially larger values than are commonly encountered for ICCs: e.g. $\rho = 0.1$ corresponds to $\psi = 0.36$ for $n = 5$ and to $\psi = 0.85$ for $n = 50$.

From (4) it is seen that estimates of β_t and θ are appropriate linear combinations of the means of the observations in each cell of the design in Table 1. An explicit formula for $\hat{\theta}$ is given in Appendix 2, where it is also shown that the coefficient matrix for $\hat{\theta}$ from (4), a_{st}, a function of ψ, is a weighted mean of the coefficient matrices for the vertical analysis, a^C_{st}, and the row-column matrix, a^{RC}_{st}, namely

$$a_{st} = \frac{(1-b)}{(1-b)+bQ}a^C_{st} + \frac{bQ}{(1-b)+bQ}a^{RC}_{st} = \frac{a^C_{st} + b(Qa^{RC}_{st} - a^C_{st})}{1 + b(Q-1)}. \quad (5)$$

Here $b = T\psi/[1 + \psi(T-1)] = Tn\rho/[1 + (Tn-1)\rho]$, a quantity which has an important role in [3] (denoted therein by R), and Q is determined by the layout of the design and is defined in Appendix 1 - for SWDs of the form in Table 1 with T periods, $Q = \frac{1}{2} + \frac{1}{2T}$. Note that $0 \le b < 1$, with $b = 0$ when $\tau^2 = 0$, and $b \to 1$ as $n\tau^2 \to \infty$. Naturally, as $b = 0$ corresponds to no cluster effect in (3), the estimator is that from the conventional model with no row effect. On the other hand, if $n\tau^2$ is large, then the between-cluster variation dominates and the smallest variance for $\hat{\theta}$ will arise from estimators that are largely based on within-cluster contrasts, and in the limit as $b \to 1$ the estimator coincides with that from the row-column design.

Equation (5) allows coefficient matrices to be computed for any T: in Table 5 we illustrate the cases with six and four periods. The pattern of coefficients is similar between the four and six period versions: the columns, but not the rows, always sum to zero; the rows sum to zero only when $\psi = 1$; when $\psi = 0$ the first and last periods are ignored. For intermediate values of ψ the row sums tend to zero as ψ increases, and observations from the first and last periods begin to contribute to $\hat{\theta}$. Notice that the observations from the first and last periods of the middle sequence (i.e. the sequence with equal replication of 0 and 1) are ignored in all analyses: this does not apply when T is odd. Note also that for all ψ the coefficient matrix has symmetry arising from the time-reversible nature of the design, namely $a_{st} = -a_{S+1-s,T+1-t}$: the matrices in Table 5 are based on a temporally reversible V, and this symmetry may not obtain if this were not the case.

3.2. Model extensions

The model (2) is suitable for the case where the nT measurements in a cluster arise from nT different individuals, each measured once. If the nT observations arise from T serial measurements on each of n individuals as would be the case in a closed cohort design [7], i.e. if the data are longitudinal, then an alternative model is needed. A suitable extension is

$$Y_{sijt} = \beta_t + D_{st}\theta + \alpha_{si} + \xi_{sij} + \epsilon_{sijt}, \quad s = 1,\ldots,S;\ i = 1,\ldots,N;\ j = 1,\ldots,n;\ ,t = 1,\ldots,T; \quad (6)$$

where ξ_{sij} is a random effect for each individual, with zero mean and variance ω^2, independent of the other random terms. A further extension is when the period effect can vary randomly between clusters, i.e. β_t in (6) is replaced by $\beta_t + \phi_{sit}$, where ϕ_{sit} are independent random variables, also independent of the other random terms and with zero mean and variance δ^2. This would be important if the way time affects the outcome differs between the clusters, for example because of differences in their location. Both of these elaborations of the model in (6) have been considered by others, e.g. [3, 11].

ψ	Coefficients SW(5)	Coefficients SW(3)
0.0	$\frac{1}{10}\begin{pmatrix} 0 & 2.0 & 1.5 & 1.0 & 0.5 & 0 \\ 0 & -0.5 & 1.5 & 1.0 & 0.5 & 0 \\ 0 & -0.5 & -1.0 & 1.0 & 0.5 & 0 \\ 0 & -0.5 & -1.0 & -1.5 & 0.5 & 0 \\ 0 & -0.5 & -1.0 & -1.5 & -2.0 & 0 \end{pmatrix}$	$\frac{1}{10}\begin{pmatrix} 0 & 5.0 & 2.5 & 0 \\ 0 & -2.5 & 2.5 & 0 \\ 0 & -2.5 & -5.0 & 0 \end{pmatrix}$
0.05	$\frac{1}{9}\begin{pmatrix} -0.2 & 1.8 & 1.3 & 0.8 & 0.3 & -0.2 \\ -0.1 & -0.6 & 1.4 & 0.9 & 0.4 & -0.1 \\ 0.0 & -0.5 & -1.0 & 1.0 & 0.5 & 0.0 \\ 0.1 & -0.4 & -0.9 & -1.4 & 0.6 & 0.1 \\ 0.2 & -0.3 & -0.8 & -1.3 & -1.8 & 0.2 \end{pmatrix}$	$\frac{1}{91}\begin{pmatrix} -4.5 & 45.5 & 20.5 & -4.5 \\ 0.0 & -25.0 & 25.0 & 0.0 \\ 4.5 & -20.5 & -45.5 & 4.5 \end{pmatrix}$
0.1	$\frac{1}{10}\begin{pmatrix} -0.4 & 2.0 & 1.4 & 0.8 & 0.2 & -0.4 \\ -0.2 & -0.8 & 1.6 & 1.0 & 0.4 & -0.2 \\ 0.0 & -0.6 & -1.2 & 1.2 & 0.6 & 0.0 \\ 0.2 & -0.4 & -1.0 & -1.6 & 0.8 & 0.2 \\ 0.4 & -0.2 & -0.8 & -1.4 & -2.0 & 0.4 \end{pmatrix}$	$\frac{1}{17}\begin{pmatrix} -1.5 & 8.5 & 3.5 & -1.5 \\ 0.0 & -5.0 & 5.0 & 0.0 \\ 1.5 & -3.5 & -8.5 & 1.5 \end{pmatrix}$
0.25	$\frac{1}{13}\begin{pmatrix} -1.0 & 2.6 & 1.7 & 0.8 & -0.1 & -1.0 \\ -0.5 & -1.4 & 2.2 & 1.3 & 0.4 & -0.5 \\ 0.0 & -0.9 & -1.8 & 1.8 & 0.9 & 0.0 \\ 0.5 & -0.4 & -1.3 & -2.2 & 1.4 & 0.5 \\ 1.0 & 0.1 & -0.8 & -1.7 & -2.6 & 1.0 \end{pmatrix}$	$\frac{1}{3}\begin{pmatrix} -0.5 & 1.5 & 0.5 & -0.5 \\ 0.0 & -1.0 & 1.0 & 0.0 \\ 0.5 & -0.5 & -1.5 & 0.5 \end{pmatrix}$
0.5	$\frac{1}{9}\begin{pmatrix} -1.0 & 1.8 & 1.1 & 0.4 & -0.3 & -1.0 \\ -0.5 & -1.2 & 1.6 & 0.9 & 0.2 & -0.5 \\ 0.0 & -0.7 & -1.4 & 1.4 & 0.7 & 0.0 \\ 0.5 & -0.2 & -0.9 & -1.6 & 1.2 & 0.5 \\ 1.0 & 0.3 & -0.4 & -1.1 & -1.8 & 1.0 \end{pmatrix}$	$\frac{1}{19}\begin{pmatrix} -4.5 & 9.5 & 2.5 & -4.5 \\ 0.0 & -7.0 & 7.0 & 0.0 \\ 4.5 & -2.5 & -9.5 & 4.5 \end{pmatrix}$
1.0	$\frac{1}{7}\begin{pmatrix} -1.0 & 1.4 & 0.8 & 0.2 & -0.4 & -1.0 \\ -0.5 & -1.1 & 1.3 & 0.7 & 0.1 & -0.5 \\ 0.0 & -0.6 & -1.2 & 1.2 & 0.6 & 0.0 \\ 0.5 & -0.1 & -0.7 & -1.3 & 1.1 & 0.5 \\ 1.0 & 0.4 & -0.2 & -0.8 & -1.4 & 1.0 \end{pmatrix}$	$\begin{pmatrix} -0.3 & 0.5 & 0.1 & -0.3 \\ 0.0 & -0.4 & 0.4 & 0.0 \\ 0.3 & -0.1 & -0.5 & 0.3 \end{pmatrix}$

Table 5. Coefficient matrices for designs SW(5) and SW(3) analysed with random effects with $V \propto (1-\psi)I_6 + \psi J_{6,6}$
(Scalings chosen so that entries to 1 d.p. are exact)

Under this model the vector of $NSnT$ responses has dispersion $I_{NS} \otimes \Omega$ where Ω is the $nT \times nT$ dispersion matrix of the responses on a cluster, and

$$\Omega = \begin{pmatrix} U & W & \cdots & W \\ W & U & \cdots & W \\ \vdots & & & \vdots \\ W & W & \cdots & U \end{pmatrix}$$

where $W = \delta^2 I_T + \tau^2 J_{T,T}$ and $U = (\nu^2 + \delta^2) I_T + (\omega^2 + \tau^2) J_{T,T}$. From this it follows that the dispersion of the T-dimensional vector of mean responses on a cluster allocated to sequence s, $\bar{Y}_{si}$ is $(n^{-1}\nu^2 + \delta^2) I_T + (n^{-1}\omega^2 + \tau^2) J_{T,T}$, i.e. the same form as for the cross-sectional design, but with τ^2 replaced by $\tau^2 + n^{-1}\omega^2$, and ν^2 by $\nu^2 + n\delta^2$. More detailed algebra shows that in the longitudinal case, and for dependencies of the form induced by the model in (6), the estimates of γ are of the form in (4), but now with $\psi = (n\tau^2 + \omega^2)/(n\tau^2 + n\delta^2 + \omega^2 + \nu^2)$, in the definition of V. Consequently the above coefficient matrices also apply to the longitudinal case, provided that the appropriate adjustment is made to ψ.

4. Variance of the treatment estimator under a variety of models

In this section an expression for the variance of $\hat{\theta}$ is derived in terms of the coefficient matrix and this is used to explore the effect of mis-specification of the dispersion structure. To be specific it is supposed that the true ICC is ρ, while that used for estimation of θ is ρ'. The notation for the matrix in (5) is extended to $a_{st}(\rho)$ in order to make explicit the correlation used in the estimation. Strictly speaking a_{st} is a function of ψ, which depends on both ρ and n, however the dependence on ρ is emphasised and that on n is suppressed, as mis-specification in the ICC ρ is more pertinent and familiar to practitioners. While the results apply to both cross-sectional and longitudinal cases, the exposition below will be in terms of the former.

If we write the $S \times T$ coefficient matrix in (5) as $a(\rho)$, where $(a(\rho))_{st} = a_{st}(\rho)$, then the estimator of θ found when the correlation is taken to be ρ' is

$$\hat{\theta}_{\rho'} = \sum_{s,t} a_{st}(\rho')(\bar{Y}_s)_t = \sum_s a_s(\rho')\bar{Y}_s,$$

where $(\bar{Y}_s)_t$ is element t of $\bar{Y}_s$ and $a_s(\rho')$ is the sth row of the coefficient matrix $a(\rho')$. The subscript ρ' has been added to $\hat{\theta}$ to emphasise the dependence of the estimator on ρ'. The true variance of $\bar{Y}_s$ is $\frac{1}{N}V$, so it follows that

$$\mathsf{var}(\hat{\theta}_{\rho'}) = \frac{1}{N}\sum_s a_s(\rho')Va_s(\rho')^T = \frac{1}{N}\mathsf{tr}(a(\rho')Va(\rho')^T), \tag{7}$$

where tr denotes the trace of the matrix. If the true variance has the equicorrelation pattern induced by (2), namely $V = \sigma^2((1-\psi)I_T + \psi J_{T,T})$, then (7) can be expressed as

$$\begin{aligned}\mathsf{var}(\hat{\theta}_{\rho'}) =& \frac{\sigma^2}{N}\{(1-\psi)\mathsf{tr}[a(\rho')a(\rho')^T] + \psi\mathsf{tr}[(a(\rho')1_T)(a(\rho')1_T)^T]\}\\ =& \frac{(\tau^2+\nu^2)}{N}\left\{\frac{(1-\rho)}{n}\mathsf{tr}[a(\rho')a(\rho')^T] + \rho\mathsf{tr}[(a(\rho')1_T)(a(\rho')1_T)^T]\right\}\end{aligned}$$

As $a(\rho')$ can be written in terms of a^{RC} and a^C, using (5), this expression becomes

$$N\mathsf{var}(\hat{\theta}_{\rho'}) = (\tau^2+\nu^2)\frac{n^{-1}(1-\rho)[(1-b')^2A_1 + b'^2Q^2A_2 + 2(1-b')b'QA_3] + \rho(1-b')^2A_4}{(1-b'+b'Q)^2}, \tag{8}$$

where

$$\begin{aligned}A_1 &= \mathsf{tr}(a^C(a^C)^T) \qquad & A_2 &= \mathsf{tr}(a^{RC}(a^{RC})^T)\\ A_3 &= \mathsf{tr}(a^{RC}(a^C)^T) \qquad & A_4 &= \mathsf{tr}((a^C1_T)(a^C1_T)^T),\end{aligned} \tag{9}$$

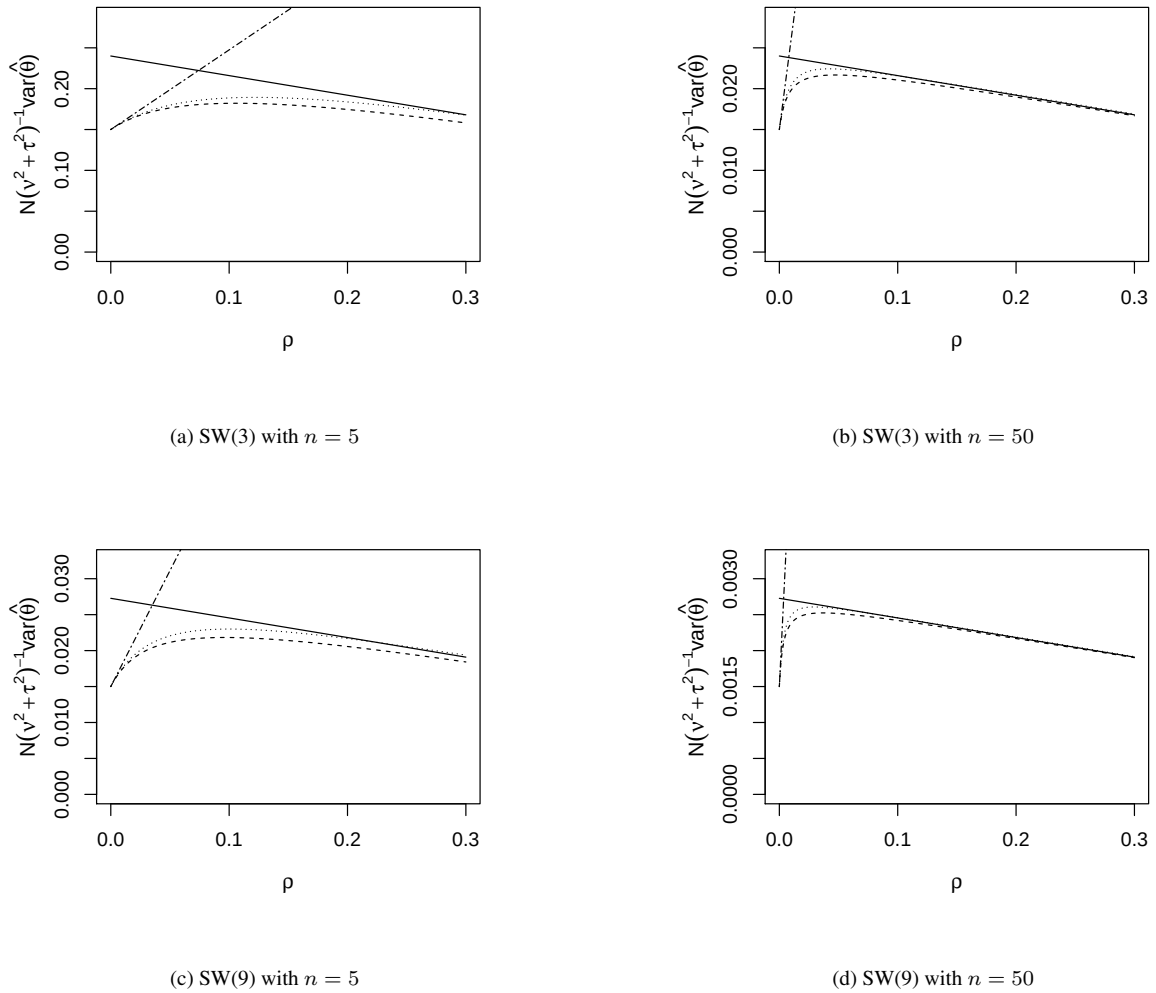
and $b' = b'(\rho') = Tn\rho'/(1+(Tn-1)\rho')$. In (8) $A_1, \ldots, A_4$ and Q are known values determined by the layout of the design, and values for classic SWDs, as in Table 1, are given in Table 6 for a range of T. Note that for the cases $\rho' = 0, 1$, $\mathsf{var}(\theta_{\rho'})$ is linearly dependent on ρ.

Analyses based on $\rho' = 0$, corresponding to the vertical analysis, on $\rho' = 1$, corresponding to the row-column analysis and on $\rho' = \rho$, being the optimal analysis with no mis-specification, are considered. So too are analyses based on perturbed values of ρ, namely $\rho' = k^{\pm 1}\rho$, for $k = 2$. For these values of ρ' Figure 1 shows the normalised variances, $(\tau^2+\nu^2)^{-1}N\mathsf{var}(\hat{\theta}_{\rho'})$, for four examples, based on SW(3), SW(9) with $n = 5, 50$. The ICC in Figure 1 ranges up to 0.3, which is considerably larger than is usually encountered in practice, but is sufficiently small that $k\rho < 1$.

Figure 1 confirms that the mixed-model estimator (4) using the true correlation for ρ' has the smallest variance. Even when ρ' is perturbed from the true correlation by a factor of 2 the effect on the variance is small: for clarity, only underestimation, which gives slightly larger variances, is shown in the figure. For $T = 4, n = 5$ the maximum over $\rho \leq 0.3$ of the ratio of the variance when $\rho' = k\rho$ to $\rho' = \rho$ is 1.060 for $k = \frac{1}{2}$, whereas the corresponding value for $k = 2$ is 1.033: for $T = 10, n = 5$ the values are 1.055 and 1.036, with slightly smaller values when $n = 50$. If $\rho = 0$ there is no clustering effect and the mixed model and vertical models coincide, while the standard row-column analysis has much

T	A_1	A_2	A_3	A_4	Q
3	2.000	3.000	2.000	2.000	0.667
4	0.750	1.200	0.750	1.125	0.625
5	0.400	0.667	0.400	0.800	0.600
6	0.250	0.429	0.250	0.625	0.583
7	0.171	0.300	0.171	0.514	0.571
8	0.125	0.222	0.125	0.438	0.562
9	0.095	0.171	0.095	0.381	0.556
10	0.075	0.136	0.075	0.337	0.550
11	0.061	0.111	0.061	0.303	0.545
12	0.050	0.092	0.050	0.275	0.542
15	0.031	0.058	0.031	0.215	0.533
20	0.017	0.032	0.017	0.158	0.525

Table 6. Values of traces from (9) and Q for standard stepped wedge designs for various $T \leq 20$



(a) SW(3) with $n = 5$

(b) SW(3) with $n = 50$

(c) SW(9) with $n = 5$

(d) SW(9) with $n = 50$

Figure 1. $N(\nu^2 + \tau^2)^{-1}\mathsf{var}(\hat{\theta}_{\rho'})$ plotted against the true value of ρ: solid line $\rho' = 1$ (row-column analysis); dashed and dotted $\rho' = 0$ (vertical analysis); dashed line $\rho' = \rho$; dotted line $\rho' = \frac{1}{2}\rho$. Designs with 4 periods (first row) and 10 periods (second row) are shown: $n = 5$ shown in first column and $n = 50$ in second column.

larger variance, because it is attempting to eliminate row effects that are, in fact, absent. With increasing clustering the variance of the vertical estimator increases markedly, and is too inefficient to be viable unless both n and ρ are small. The efficiency of the row-column estimator increases with increasing ρ; for $\rho > 0.2$ its variance is less than 10% ($T = 4$) and 6% ($T = 10$) larger than the optimal value when $n = 5$, and only 1% and 0.6% larger when $n = 50$.

5. Efficient vertical analyses

The so-called vertical analysis is based on the coefficient matrix a^C: the 'vertical' aspect is that the columns alone sum to zero, so providing contrasts made between randomized groups in each period of the design. The first and last columns vanish because there is no variation between treatments in these columns, so $\hat{\theta}$ based on a^C is a weighted combination of the $T-2$ non-trivial column-wise estimators. The estimator using a^C weights the column-wise estimators inversely by their variance and is optimal in the case when the elements in a row are independent.

$$\begin{pmatrix} 0 & 0.2 & 0.15 & 0.10 & 0.05 & 0 \\ 0 & -0.05 & 0.15 & 0.10 & 0.05 & 0 \\ 0 & -0.05 & -0.10 & 0.10 & 0.05 & 0 \\ 0 & -0.05 & -0.10 & -0.15 & 0.05 & 0 \\ 0 & -0.05 & -0.10 & -0.15 & -0.20 & 0 \end{pmatrix}$$

(a) a^C for six-period design

$$\begin{pmatrix} 0 & 0.2\lambda_1 & 0.15\lambda_2 & 0.10\lambda_3 & 0.05\lambda_4 & 0 \\ 0 & -0.05\lambda_1 & 0.15\lambda_2 & 0.10\lambda_3 & 0.05\lambda_4 & 0 \\ 0 & -0.05\lambda_1 & -0.10\lambda_2 & 0.10\lambda_3 & 0.05\lambda_4 & 0 \\ 0 & -0.05\lambda_1 & -0.10\lambda_2 & -0.15\lambda_3 & 0.05\lambda_4 & 0 \\ 0 & -0.05\lambda_1 & -0.10\lambda_2 & -0.15\lambda_3 & -0.20\lambda_4 & 0 \end{pmatrix}$$

(b) Modified vertical analysis

Table 7. Coefficient matrices for vertical analyses of designs SW(5): λ_i chosen to ensure unbiassedness and minimum variance when there is within-row correlation

In practice it is not possible to exclude within-row dependence in the outcomes. An alternative approach to providing a 'vertical' analysis is to use the coefficient matrix in table 7(b), where $\lambda_1, \ldots, \lambda_{T-2}$ are chosen to minimise the variance of $\hat{\theta}$, while maintaining unbiassed estimation of θ. In the following it will be convenient to define $\tilde{a}$ to be the $S \times (T-2)$-dimensional matrix obtained by omitting the first and last columns from a^C. Also, suppose that $\tilde{X}$ denotes the $S \times (T-2)$-dimensional matrix of 0s and 1s showing the SWD, as in Table 1, again with the first and last columns omitted. If the column weights $\lambda_1, \ldots, \lambda_{T-2}$ are written as a diagonal matrix Λ, then the revised estimator, $\hat{\theta}_\Lambda$ is that obtained by using $\tilde{a}\Lambda$ as the coefficient matrix. Also

$$\mathsf{var}(\hat{\theta}_\Lambda) = \frac{1}{N}\mathsf{tr}(\tilde{a}\Lambda\tilde{V}\Lambda\tilde{a}^T) = \frac{1}{N}\mathsf{tr}(\Lambda\tilde{V}\Lambda B)$$

and for unbiassedness

$$1 = \mathsf{tr}(\tilde{a}\Lambda\tilde{X}^T) = \mathsf{tr}(\Lambda\tilde{X}^T\tilde{a}),$$

where $B = \tilde{a}^T\tilde{a}$ and $\tilde{V}$ is the dispersion matrix for the cluster means in periods 2 to $T-1$, i.e. $\tilde{V}$ is V with the first & last rows and the first & last columns omitted. To minimise $\mathsf{var}(\hat{\theta}_\Lambda)$, introduce the Lagrange multiplier μ and Lagrangian

$$L = \mathsf{tr}(\Lambda\tilde{V}\Lambda B) - \mu(\mathsf{tr}(\Lambda\tilde{X}^T\tilde{a}) - 1)$$

and

$$\frac{\partial L}{\partial \lambda_k} = 2\mathsf{tr}(e_k\tilde{V}\Lambda B) - \mu\mathsf{tr}(e_k\tilde{X}^T\tilde{a}), \quad k = 1, \ldots, T-2$$

where e_k is a $(T-2) \times (T-2)$ matrix of zeroes except that element (k,k) is 1. The first term on the LHS is $2\sum_j \tilde{v}_{kj}b_{kj}\lambda_j$ and the second term is the kth diagonal element of $\tilde{X}^T\tilde{a}$, which we collect together as a $(T-2)$-dimensional vector d.

Writing $\tilde{V} \circ B$ for the Hadamard, or element-wise product of $\tilde{V}$ and B, the optimal λ is

$$\lambda_{\text{opt}} = \frac{(\tilde{V} \circ B)^{-1} d}{d^T (\tilde{V} \circ B)^{-1} d}.$$

If $\tilde{V}$ is of the form $\omega_1 I_{T-2} + \omega_2 J_{T-2,T-2}$, then $\tilde{V} \circ B = \omega_1 \mathsf{diag}(B) + \omega_2 B$, where $\mathsf{diag}(\cdot)$ denotes the diagonal elements of its argument. Using this form it can be shown that $(\omega_1 \mathsf{diag}(B) + \omega_2 B) 1_{T-2} \propto d$, in other words for equicorrelation dependence, the optimum vertical analysis sets all $\lambda_i = 1$, i.e. to use the vertical analysis derived in Section 2 based on a^C. This does not have to apply if $\tilde{V}$ takes another form: for example with $T = 6$ and V having an autocorrelation structure with $V_{ij} = \rho^{|i-j|}$, then for $\rho = \frac{1}{2}$ the optimal λs are 1.108, 0.929, 0.929, 1.108. However, in this case the distinction is slight, because the efficiency of the equally weighted analysis is around 99.6%.

The above indicates that for the equicorrelation structure the equally weighted vertical analysis, shown to be generally very inefficient in Section 4 is, nonetheless, the best of the extended vertical analyses exemplified in Table 7(b). It is only slightly suboptimal amongst vertical analyses when an autoregressive dispersion model is assumed, which demonstrates that even the best extended vertical analysis is highly inefficient for some plausible dispersion structures.

6. Extended SWDs with multiple treatments

6.1. Row-column and vertical analyses

Although the classic form in Table 1 dominates the literature on SWDs, numerous variants are possible, several of which fit into the framework of the present paper. The extensions considered in this section are to designs where a standard treatment, 0, is compared with two new treatments, denoted by 1 and 2. There are several types of such designs, each potentially of use in a variety of settings. Lyons and colleagues [12] present and categorise various extensions of this type, and cite four trials where these extensions either have been used or might profitably have been used [13, 14, 15, 16].

	Time		
Sequence	1	2	3
1	0	1	1
2	0	0	1
3	0	2	2
4	0	0	2

Table 8. Concurrent stepped wedge design with $T = 3$ periods, $S = 4$ sequences comparing a standard treatment 0 with two new treatments 1 & 2.

The first extension is the *concurrent* design, shown in Table 8 for three periods, which is essentially two separate SW(2) designs each testing a different new intervention. Although these assessments could be made in two separate classic SW(2) designs, it may be sensible to combine them as each new treatment is compared with the same standard treatment, and if it is reasonable to believe that the same period effects extend across the two component SW(2) designs. In this case a suitable extension to the model in (1) is

$$Y_{st} = \alpha_s + \beta_t + D_{1st}\theta_1 + D_{2st}\theta_2 + \eta_{st} \tag{10}$$

where $D_{kst} = 1$ if treatment $k \in \{1, 2\}$ is allocated to cell (s, t), and 0 otherwise: here θ_k is the effect of treatment k relative to treatment 0.

Two further extensions considered in [12] are the *replacement* and *supplementation* designs, shown in Table 9 for the case $T = 6$. It is suggested [12] that the replacement design would be useful when there are two interventions to assess but

	Time					
Sequence	1	2	3	4	5	6
1	0	1	1	2	2	2
2	0	0	1	1	2	2
3	0	0	0	1	1	2

(a) Replacement design for $T = 6$

	Time					
Sequence	1	2	3	4	5	6
1	0	1	1	1+2	1+2	1+2
2	0	0	1	1	1+2	1+2
3	0	0	0	1	1	1+2

(b) Supplementation design for $T = 6$

Table 9. Replacement and Supplementation designs for comparing two new treatments with a standard

which cannot be applied at the same time. It is conceded that there must be no carryover from the periods when intervention 1 is used into those when intervention 2 is employed. In many service implementation studies the new arrangements may take some time to become familiar, so a further use of the replacement design could be practical way to allow for this effect, by conceiving of treatment 2 as a mature version of treatment 1, as has been previously suggested [10, 17].

The supplementation design would be suitable when a second intervention requires that the first is present, or where the first cannot be removed once it has been implemented. Mathematically the design is largely the same as the replacement design, as from this point of view the difference between the designs amounts to a re-parametrization of the treatment effect. In the supplementation design $\hat{\theta}_2$ is $\hat{\theta}_2 - \hat{\theta}_1$ in the replacement design.

Other extensions are mentioned in [12], such as a factorial design; the approach described below can also be applied to these designs and they will be explored elsewhere.

The row-column and vertical analyses described in Section 2 can be applied to these extended designs if the model in (10) is adopted. The estimator of $(\theta_1, \theta_2)^T$ is derived in (A3) and, as for the two-treatment case, can be expressed in terms of coefficient matrices. These can be written in terms of the coefficient matrices for two sub-designs, $D1$ and $D2$ of a three treatment design D: cell (s, t) of $D1(D2)$ is 1 if treatment 1(2) is applied in that cell and 0 otherwise. If the coefficient matrices for $D1$ and $D2$, regarded as two-treatment designs, are $a(1)$ and $a(2)$ respectively then the coefficient matrices for $(\hat{\theta}_1, \hat{\theta}_2)$ from D can be written as

$$\begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix} = \tilde{M}^{-1} \begin{bmatrix} a(1) \\ a(2) \end{bmatrix}.$$

As $a(1), a(2)$ are $S \times T$ matrices, the [] are a reminder that this equation must be interpreted formally, namely as the implied linear combination of the coefficient matrices: cf. (A4). The matrix $\tilde{M}$ is

$$\begin{pmatrix} 1 & r_1 \\ r_2 & 1 \end{pmatrix},$$

so, for example, the coefficient matrix for $\hat{\theta}_1$ is $[a(1) - r_1 a(2)]/(1 - r_1 r_2)$. The elements r_1 and r_2 differ depending on whether the row-column or the vertical analysis is used. In addition, the dispersion matrix of the estimators can also be found as:

$$\mathsf{var} \begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix} = \frac{\sigma_\eta^2}{N} \tilde{M}^{-1} \begin{pmatrix} \mathsf{tr}(a(1)a(1)^T) & \mathsf{tr}(a(1)a(2)^T) \\ \mathsf{tr}(a(1)a(2)^T) & \mathsf{tr}(a(2)a(2)^T) \end{pmatrix} \tilde{M}^{-1T}. \tag{11}$$

For the concurrent design the coefficients and dispersion matrices are shown in Table 10 for both the row-column and vertical analyses. The variance of $\hat{\theta}_1$ from a design which comprised just the first two rows from Table 8 would be $3\sigma_\eta^2$ for the row-column analysis and $2\sigma_\eta^2$ for the vertical analysis. Consequently, while an investigator assessing two interventions against the same standard treatment might use a row-column analysis of two separate SW(2) designs, obtaining $\mathsf{var}(\hat{\theta}_1) = v$, the same resources could give a variance of $\frac{15}{24}v$ if the concurrent design in Table 8 were used. In this case the investigator would have to be persuaded that the period effects β_t could be assumed to apply to the whole design, because it is the availability of information on these parameters from all the rows in Table 8 which gives rise to

the more precise treatment estimator.

	(a)	(b)
Coefficient matrices	$\frac{1}{8}\begin{pmatrix} -4 & 5 & -1 \\ 1 & -5 & 4 \\ 0 & 3 & -3 \\ 3 & -3 & 0 \end{pmatrix}$	$\frac{1}{6}\begin{pmatrix} 0 & 4 & 1 \\ 0 & -3 & 1 \\ 0 & 2 & -1 \\ 0 & -3 & -1 \end{pmatrix}$
$\sigma_\eta^{-2}\mathsf{var}\begin{pmatrix}\hat\theta_1\\ \hat\theta_2\end{pmatrix}$	$\frac{1}{8}\begin{pmatrix} 15 & 9 \\ 9 & 15 \end{pmatrix}$	$\frac{1}{6}\begin{pmatrix} 7 & 5 \\ 5 & 7 \end{pmatrix}$

Table 10. Coefficient matrices for $\hat\theta_1$ for the concurrent design in Table 8: (a) row-column and (b) vertical analyses. The coefficient matrices for θ_2 are the same but with the first two rows interchanged with the second two rows. The dispersion matrices for $(\hat\theta_1, \hat\theta_2)^T$ for these models are shown below the coefficient matrices

The coefficient and dispersion matrices for the Replacement design (Table 9(a)) are shown in Table 11. As remarked in [12], while the concurrent design gives $\hat\theta_1$ and $\hat\theta_2$ that have equal precision, the replacement design gives an estimate of $\hat\theta_2$ which is notably less precise than $\hat\theta_1$. Although the two treatments are equally replicated, in the replacement design treatment 2 is more heavily confounded with period effects than treatment 1. The estimates of the two treatment effects are highly correlated in both row-column and vertical analyses. It is worth noting that the coefficient matrices in Table 11 contain a number of zeroes, corresponding to observations that will be ignored in the analysis. This is more marked for the vertical analysis but nonetheless there are a number of observations which contribute nothing even in the row-column analysis.

Replacement Design

Sequence 1	0	1	1	2	2	2
Sequence 2	0	0	1	1	2	2
Sequence 3	0	0	0	1	1	2

	(a)	(b)
$\hat\theta_1$	$\frac{1}{8}\begin{pmatrix} -2 & 4 & 1 & 0 & -1 & -2 \\ 0 & -3 & 3 & -1 & 1 & 0 \\ 2 & -1 & -4 & 1 & 0 & 2 \end{pmatrix}$	$\frac{1}{4}\begin{pmatrix} 0 & 2 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & -2 & 0 & 0 & 0 \end{pmatrix}$
$\hat\theta_2$	$\frac{1}{2}\begin{pmatrix} -1 & 1 & 0 & 1 & 0 & -1 \\ 0 & -1 & 1 & -1 & 1 & 0 \\ 1 & 0 & -1 & 0 & -1 & 1 \end{pmatrix}$	$\frac{1}{4}\begin{pmatrix} 0 & 2 & 1 & 2 & 1 & 0 \\ 0 & -1 & 1 & -1 & 1 & 0 \\ 0 & -1 & -2 & -1 & -2 & 0 \end{pmatrix}$
$\sigma_\eta^{-2}\mathsf{var}\begin{pmatrix}\hat\theta_1\\ \hat\theta_2\end{pmatrix}$	$\frac{1}{8}\begin{pmatrix} 9 & 12 \\ 12 & 24 \end{pmatrix}$ $\frac{1}{8}\begin{pmatrix} \mathit{9} & \mathit{3} \\ \mathit{3} & \mathit{9} \end{pmatrix}$	$\frac{1}{4}\begin{pmatrix} 3 & 3 \\ 3 & 6 \end{pmatrix}$ $\frac{1}{4}\begin{pmatrix} \mathit{3} & \mathit{0} \\ \mathit{0} & \mathit{3} \end{pmatrix}$

Table 11. Coefficient and dispersion matrices for the row-column (a) and vertical (b) analyses of the six-period replacement SW design: note italicised disperson matrices in last row relate to the Supplementation design.

The corresponding results for the Supplementation design, Table 9(b), are not shown because they are closely related to those for the Replacement design. In an obvious notation, $\theta_1^{\mathrm{Supp}} = \theta_1^{\mathrm{Rep}}$ and $\theta_2^{\mathrm{Supp}} = \theta_2^{\mathrm{Rep}} - \theta_1^{\mathrm{Rep}}$, so if the coefficient matrices from the Replacement design for θ_1 and θ_2 are, respectively a_1^{Rep} and a_2^{Rep}, then the corresponding matrices for the Supplementation design are $a_1^{\mathrm{Supp}} = a_1^{\mathrm{Rep}}$ and $a_2^{\mathrm{Supp}} = a_2^{\mathrm{Rep}} - a_1^{\mathrm{Rep}}$. Moreover this connection between the parameters allows the variances for the supplementation design to be found, as shown in italics in Table 11: note that in this parameterisation θ_1 and θ_2 are estimated with equal precision and are markedly less correlated than for the Replacement design.

6.2. Random effects analyses

The random effects model described in Section 3.1 carries over to the extended designs in a natural way. If the analysis of an extended design D uses the correlation ρ' and the coefficient matrices for the sub-designs $D1$ and $D2$ are $a(\rho', 1)$ and

$a(\rho', 2)$ respectively, then the estimators for the treatment effects are

$$\begin{pmatrix} \hat{\theta}_{\rho' 1} \\ \hat{\theta}_{\rho' 2} \end{pmatrix} = \tilde{M}^{-1} \begin{bmatrix} a(\rho', 1) \\ a(\rho', 2) \end{bmatrix} \tag{12}$$

where now

$$\tilde{M} = \begin{pmatrix} 1 & \tilde{Q}_1 \frac{1-b'+b'Q_{12}}{1-b'+b'Q_1} \\ \tilde{Q}_2 \frac{1-b'+b'Q_{12}}{1-b'+b'Q_2} & 1 \end{pmatrix}$$

and, as before, $b' = Tn\rho'/(1 + (Tn-1)\rho')$ and the Qs are defined at the end of Appendix 1. The dispersion matrix of $(\hat{\theta}_{\rho' 1}, \hat{\theta}_{\rho' 2})^T$ also carries over from the uncorrelated case so, using the new form for $\tilde{M}$, we have from (11) that

$$\mathsf{var} \begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix} = \frac{1}{N} \tilde{M}^{-1} \begin{pmatrix} \mathsf{tr}(a(\rho', 1) V a(\rho', 1)^T) & \mathsf{tr}(a(\rho', 1) V a(\rho', 2)^T) \\ \mathsf{tr}(a(\rho', 1) V a(\rho', 2)^T) & \mathsf{tr}(a(\rho', 2) V a(\rho', 2)^T) \end{pmatrix} \tilde{M}^{-1T},$$

where V is the true variance, as used in (7).

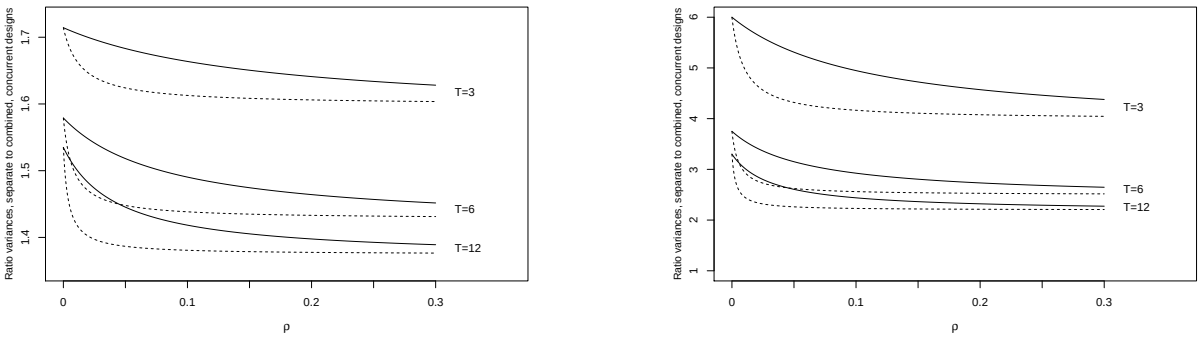
ψ'	Coefficients combined	Coefficients separate
0.0	$\frac{1}{6}\begin{pmatrix} 0 & 4 & 1 \\ 0 & -3 & 1 \\ 0 & 2 & -1 \\ 0 & -3 & -1 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 0 \\ 0 & -1 & 0 \end{pmatrix}$
0.05	$\frac{1}{1281}\begin{pmatrix} -51.5 & 850.5 & 179.5 \\ 20.0 & -651.0 & 251.0 \\ -9.5 & 430.5 & -240.5 \\ 41.0 & -630.0 & -190.0 \end{pmatrix}$	$\frac{1}{21}\begin{pmatrix} -1 & 21 & -1 \\ 1 & -21 & 1 \end{pmatrix}$
0.1	$\frac{1}{341}\begin{pmatrix} -26.5 & 225.5 & 39.5 \\ 10.0 & -176.0 & 76.0 \\ -4.5 & 115.5 & -70.5 \\ 21.0 & -165.0 & -45.0 \end{pmatrix}$	$\frac{1}{11}\begin{pmatrix} -1 & 11 & -1 \\ 1 & -11 & 1 \end{pmatrix}$
0.25	$\frac{1}{13}\begin{pmatrix} -2.3 & 8.5 & 0.7 \\ 0.8 & -7.0 & 3.8 \\ -0.3 & 4.5 & -3.3 \\ 1.8 & -6.0 & -1.2 \end{pmatrix}$	$\begin{pmatrix} -0.2 & 1 & -0.2 \\ 0.2 & -1 & 0.2 \end{pmatrix}$
0.5	$\frac{1}{21}\begin{pmatrix} -6.5 & 13.5 & -0.5 \\ 2.0 & -12.0 & 8.0 \\ -0.5 & 7.5 & -6.5 \\ 5.0 & -9.0 & -1.0 \end{pmatrix}$	$\frac{1}{3}\begin{pmatrix} -1 & 3 & -1 \\ 1 & -3 & 1 \end{pmatrix}$
1.0	$\frac{1}{4}\begin{pmatrix} -2.0 & 2.5 & -0.5 \\ 0.5 & -2.5 & 2.0 \\ 0.0 & 1.5 & -1.5 \\ 1.5 & -1.5 & 0.0 \end{pmatrix}$	$\begin{pmatrix} -0.5 & 1 & -0.5 \\ -0.5 & -1 & 0.5 \end{pmatrix}$

Table 12. Coefficient matrices for $\hat{\theta}_1$ for the three-period concurrent design in Table 8 using random effects model (left hand column). Although the matrices, $a(\rho', 1), a(\rho', 2)$ are written in terms of ρ' for consistency with Section 4, they are functions of ρ' and n through $\psi' = n\rho'/[(1-\rho') + n\rho']$, which is used to index this table. Coefficients for SW(2) (right hand column) shown for comparison.(Scalings chosen so that entries to 1 d.p. are exact)

If an investigator wishes to assess three treatments using the design D in Table 8, then the treatment estimator can be based on an analysis of the whole design, using the coefficients in the left-hand column of Table 12. An alternative approach would be to analyse the component designs separately, using the coefficients in the right-hand column of Table 12. This might be adopted if the analyst is unsure if the period effects can be taken to be the same across $D1$ and $D2$. As

with the classic design, information from the first period is ignored when $\rho' = \psi' = 0$, but in the combined analysis, it is no longer necessary to ignore information from period 3. Again, when $\rho' = \psi' = 1$ the coefficients within a sequence will always sum to 0. For intermediate values the pattern of coefficients changes smoothly from one extreme analysis to the other.

Assuming that equal numbers are allocated to all the sequences in D, the variances of $\hat{\theta}_1$ from the combined and separate analyses can be compared and the ratio of these two quantities is plotted in Figure 2(a) for the design in Table 8. With this formulation the effect of mis-specification of ρ could be investigated, but this is not considered in Figure 2 where ρ' is assumed to be equal to the true correlation ρ. The design in Table 8 combines two copies of SW(2), and Figure 2 also shows the results for the analogous designs based on SW(5) and SW(11). The advantage of using the combined design reduces with increasing ρ but even for ρ close to 0.3 the separate analysis has a variance between 1.4 and 1.6 times larger, with the larger advantage occurring for shorter designs.



(a) Variances of $\hat{\theta}_1$ (or $\hat{\theta}_2$)

(b) Variances of $\hat{\theta}_1 - \hat{\theta}_2$

Figure 2. Ratio of variances separate:combined, for designs with 3, 6 and 12 periods: solid lines, $n = 5$ for solid and $n = 50$ for dashed lines

If there is interest in estimating $\theta_1 - \theta_2$ then the combined analysis analysis offers a much more efficient estimator than simply taking the difference of the estimators from the separate analyses. The corresponding variance ratios for $\hat{\theta}_1 - \hat{\theta}_2$ are show in Figure 2(b): while the overall pattern is similar, the ratios indicate a much larger advantage for the combined analysis.

7. Discussion

In stepped wedge trials the usual precepts of design theory have to be sacrificed to meet the practical constraints of the applications where they are typically used. Nevertheless a SWD is still an experimental design and some of its properties can best be understood by viewing it in this light. In particular, because of the marked non-orthogonality between the period effects and the treatment effect inherent in a SWD, it is useful for the triallist to understand how the observations are combined to provide a treatment estimator for each of the several forms of analysis that are available. The coefficient matrices derived in Sections 2, 3.1 and 6 allow the user to see how the information in the different cells of the table are combined to provide $\hat{\theta}$: Kershner and Federer [18] provided similar information for two-treatment crossover designs. The treatment estimators obtained from most designed experiments apply quite similar weights to the observations, as shown in Table 3(b). This is not the case for SWDs: several of the coefficient matrices shown in the present paper reveal that many observations receive relatively low weight, or have coefficents with counterintuitive signs, and some observations are ignored altogether. Of course, more efficient and balanced designs, such as NrO, are not alternatives to SWDs because

they do not fulfil some of the requirements such as treatments changing from 0 to 1 only once in a sequence, which in practice corresponds to a new intervention not being able to be withdrawn once it has been implemented. Nevertheless, it is important for users of SWDs to be aware of the statistical consequences of their choice of design.

The coefficient matrices presented in this paper assume that the number of clusters in a sequence, N, and the number of observations in each period within each cluster, n are constant. In practice this will not be the case, but i) departures will often be small, and ii) this assumption provides coefficients which are easily assimilated, thereby giving greater insight into the way SWDs provide information on the treatment effects.

For designs comparing two treatments the coefficients are often simpler for the vertical analysis. However, as Figure 1 shows, the vertical analysis produces the most precise estimator only if there is no variation between clusters. As the contribution of the between-cluster variation to the total variation in the cluster means $\bar{Y}_{is}$ increases, the vertical analysis rapidly produces a very much less precise estimator than a mixed-effects or row-column analysis. Substantial values of ψ are likely to be a feature of many SWDs, especially those with larger values for n, so the efficient use of these designs will require the analyst to accept the unbalanced contrasts, such as those shown in Table 5.

The results in Sections 3 and 4 are largely based on the form for V induced by (3), although similar results would obtain for the extensions described in Section 3.2. However there may be good reason to consider other forms for V, such as ones where cluster effects are less correlated the further they are separated in time. Such V could be used in (4) although analogues of the subsequent algebraic results would only be available if the revised V possessed an explicit inverse. We hope to report on the application of autoregressive processes to these designs in the near future.

The extended designs proposed in [12] and discussed in Section 6 represent quite a new line of work but may offer substantial advantages in certain contexts. In particular the concurrent design, running two SWDs in parallel, allows period effects to be eliminated more efficiently, and, if relevant, the two new treatments can be compared much more efficiently. The Replacement design, Table 9(a), allows a second new treatment to follow the first. However, if it is not imperative that all clusters end the trial receiving the second treatment, then other forms of the design may have merit. For example, if the middle sequence in Table 9(a) were permitted to be $0, 0, 2, 2, 1, 1$, then $\mathsf{var}(\hat{\theta}_1)$ decreases by 8% and $\mathsf{var}(\hat{\theta}_2)$ by 49%; the large reduction for θ_2 is because in the new design there is much less confounding of the information on θ_2 with the period effect. If a need for designs of this kind were established, then further exploration of their precise form could pay substantial dividends.

Other forms of SWD have also been discussed, some of which fit into the framework derived above, such as hybrid designs [3], where the classic SWD is supplemented with a number of sequences in which only treatment 0, or only treatment 1 is used, so it is a mixture of a SWD with a standard cluster randomized trial. In the classic SWD the number of periods is one more than the number of sequences, because the changeover from standard to new treatment advances by one period every sequence. However, this pattern can be modified, as shown in [3] and provided observations are made in every period and every sequence, the coefficient matrices can be derived using the methods in Section 2. Other patterns, such as periods where no observations are taken in some sequences, feature in some designs[19], where such interludes are used to accommodate training sessions; the methods in Sections 2 and 3 would need some adaptations to deal with such cases.

The possibility that responses may vary over time, as well as because of changes in treatment, means that a model which allows for time effects is essential. The same is true of crossover trials, where the prudent analyst may wish to include period effects. The usual way that time is taken into account is through the term β_t in (2). Non-standard approaches can be used if a specially tailored model can be identified for a particular application [20], but this is the exception not the rule. In crossover trials empirically determined departures from the standard model are seldom encountered because crossover studies usually recruit too few subjects to permit a secure exploration of alternative models. The position with SWDs may be somewhat different, as these trials will often be much larger and the data might suggest models that are better than that in (2). Allowing for time by replacing the categorical variables with a simple trend has been mentioned [12] as one possibility, although other forms might be contemplated. Moreover, it may be plausible in some applications to assume

that the effect of time varies between clusters, as described in Section 3.2. Obviously there would be dangers in using the data from a particular study to determine its analysis, but as experience accrues in a particular field, useful information could emerge.

In the early development of crossover trials it was usually assumed that treatment effects might carry over into the immediately succeeding period, in a way described by a simple model [21]. More recently there has been substantial criticism of the applicability of such a model [6, 22], and approaches in which carryover effects are excluded by non-statistical methods, such as washout periods, have found favour. This is another aspect where the small size of crossover trials has inhibited attempts to identify alternative models by empirical exploration of data.

In classic SWDs something directly analogous to a carryover cannot arise because the new treatment is never followed by any other treatment (although issues may arise for the designs in Section 6). What can arise is that the effect of the new treatment changes in the time shortly after the changeover in treatments. Such an effect might happen if, for example, staff delivering a new treatment become more proficient as they become more accustomed to a new way of working [7]. Wash-in periods in SWDs, placed just after the change in treatment when staff get used to the new treatment but the data are not analysed, is a design device to mitigate this problem, analogous to the well-established use of wash-out periods in crossovers. An alternative approach [10, 23, 17] is to use (2) but with values of D_{st} other that 0 or 1. For example fractional D_{st} might be specified just after the treatment changes over, perhaps increasing to 1 in steps defined by the analyst. In such circumstances much of the work in the present paper would carry over with only minor adjustments. For example, the matrix a_{st}^{RC} is still more or less as implied by (A1), namely

$$a_{st}^{RC} = \frac{STD_{st} - SR_{1s} - TC_{1t} + G_1}{ST\tilde{G}_1 - SR_1^T R_1 - TC_1^T C_1 + G_1^2},$$

with the obvious caveats that $R_1(C_1)$ are the row(column) sums of the D_{st} and $G_1 = \sum_{s,t} D_{st}$. The main change is that the first G_1 on the denominator of (A1) changes to $\tilde{G}_1 = \sum_{s,t} D_{st}^2$. Clearly G_1 and $\tilde{G}_1$ coincide when $D_{st} \in \{0, 1\}$ but not otherwise. Likewise, the formula for $\mathsf{var}(\hat{\theta})$ in [10] can be used for general X_{ij} given a similar amendment, namely that

$$\mathsf{var}(\hat{\theta}) = \frac{I\sigma^2(\sigma^2 + T\tau^2)}{(I\tilde{U} - W)\sigma^2 + (U^2 + IT\tilde{U} - TW - IV)\tau^2}$$

where $\tilde{U} = \sum_{i,j} X_{ij}^2$.

In summary, this paper describes the stepped wedge design in terms of experimental design concepts, which has enabled the form and nature of the treatment effet estimator to be made explicit and its properties more clearly understood.

Appendix 1: derivation of coefficient matrix for row-column design

Two-treatment case

Starting from a row-column design with S rows and T columns comparing two treatments, and with a single continuous outcome y_{st} in the sth row and tth column, then the model in (1) can be written in matrix form as $y = X\zeta + \eta$, where the parameters are written in vector form as $\zeta^T = (\alpha^T, \beta^T, \theta)$ and the design matrix is

$$X = \left(I_S \otimes 1_T \;\middle|\; 1_S \otimes I_T \;\middle|\; D_1 \right),$$

where D_1 is an ST-dimensional vector of 0s and 1s indicating treatment allocation and 1_m denotes an m-dimensional vector of 1s. The least squares estimator of ζ is $(X^TX)^-X^Ty$, where A^- is a g-inverse of A: in order to obtain an expression for the estimator of θ, we need to consider partitions of $(X^TX)^-$ and X^Ty. If we denote the $m \times n$ matrix of 1s by $J_{m,n}$, then

$$X^TX = \left(\begin{array}{cc|c} TI_S & J_{S,T} & R_1 \\ J_{T,S} & SI_T & C_1 \\ \hline R_1^T & C_1^T & G_1 \end{array}\right),$$

where R_1 (C_1) is an S (T)-dimensional vector giving the number of times treatment 1 appears in each row (column), and G_1 is the number of times treatment 1 appears in the whole design. One of the steps in finding an expression for $\hat{\theta}$ is to invert this matrix and this is facilitated by noting that a g-inverse of the leading $(S+T) \times (S+T)$ sub-matrix of X^TX is

$$\begin{pmatrix} TI_S & J_{S,T} \\ J_{S,T}^T & SI_T \end{pmatrix}^- = \begin{pmatrix} \frac{1}{T}I_S - aJ_{S,S} & \frac{1}{(S+T)^2}J_{S,T} \\ \frac{1}{(S+T)^2}J_{S,T}^T & \frac{1}{S}I_T - bJ_{T,T} \end{pmatrix}$$

where

$$a = \frac{2T+S}{T(S+T)^2} \;\text{ and }\; b = \frac{2S+T}{S(S+T)^2}.$$

Application of standard results on partitioned matrices then shows that the estimate of the treatment effect from the row-column model is

$$\hat{\theta} = \frac{STy_{(1)} - SR_1^Ty_R - TC_1^Ty_C + G_1y_G}{STG_1 - SR_1^TR_1 - TC_1^TC_1 + G_1^2}, \tag{A1}$$

where $y_{(1)}$ is the sum of the y_{st} which have received treatment 1, y_R is the vector of row sums of the y_{st}, i.e. the sth element of y_R is $\sum_{t=1}^{T} y_{st}$ and y_C is the vector of column sums, and y_G is the sum of all the y_{st}. From this it can be seen that the coefficient matrix a_{st}^{RC} is proportional to $STD_{st} - SR_{1s} - TC_{1t} + G_1$.

The corresponding expression when the row effect is omitted from the model is

$$\hat{\theta} = \frac{Sy_{(1)} - C_1^Ty_C}{SG_1 - C_1^TC_1}, \tag{A2}$$

indicating that a_{st}^C is now proportional to $SD_{st} - C_{1t}$. For the case of three treatments, and for the case of random cluster effects, it is useful to make the following definitions: d_{RC} as the denominator of (A1), d_C to be T times the denominator of (A2) and $Q = d_{RC}/d_C$.

Three-treatment case

For row-column designs which compare a standard treatment with two new treatments, the design matrix can be extended to

$$X = \left(I_S \otimes 1_T \;\middle|\; 1_S \otimes I_T \;\middle|\; D_1 \;\middle|\; D_2 \right),$$

where the effects are $\hat{\theta}_1$ and $\hat{\theta}_2$ (no interaction is considered) and the ST-dimensional vectors D_1 and D_2 describe the allocation of treatments 1 and 2 in the design. Suppose R_2, C_2 and G_2 are defined as the analogues of R_1, C_1 and G_1 for treatment 2, and G_{12} is the number of cells where treatments 1 and 2 are both allocated. The estimator of $(\theta_1, \theta_2)^T$ is

$$\begin{pmatrix}\hat{\theta}_1\\ \hat{\theta}_2\end{pmatrix} = M^{-1}\begin{pmatrix}STy_{(1)} - SR_1^T y_R - TC_1^T y_C + G_1 y_G\\ STy_{(2)} - SR_2^T y_R - TC_2^T y_C + G_2 y_G\end{pmatrix} \tag{A3}$$

where

$$M = ST\begin{pmatrix}G_1 & G_{12}\\ G_{12} & G_2\end{pmatrix} + \begin{pmatrix}G_1^2 & G_1G_2\\ G_1G_2 & G_2^2\end{pmatrix} - S\begin{pmatrix}R_1^TR_1 & R_1^TR_2\\ R_1^TR_2 & R_2^TR_2\end{pmatrix} - T\begin{pmatrix}C_1^TC_1 & C_1^TC_2\\ C_1^TC_2 & C_2^TC_2\end{pmatrix},$$

and $y_{(2)}$ is the sum of observations receiving treatment 2. The coefficients a_{st}^{RC} can then be identified using the approach used for the two-treatment case. The result for three treatments when the row effect is omitted is

$$\hat{\theta} = \left[\begin{pmatrix}G_1 & G_{12}\\ G_{12} & G_2\end{pmatrix} - \frac{1}{S}\begin{pmatrix}C_1^TC_1 & C_2^TC_1\\ C_2^TC_1 & C_2^TC_2\end{pmatrix}\right]^{-1}\left[\begin{pmatrix}y_1\\ y_2\end{pmatrix} - \frac{1}{S}\begin{pmatrix}C_1^Ty_C\\ C_2^Ty_C\end{pmatrix}\right].$$

If D is a design with three treatments, then the coefficient matrices for $\hat{\theta}_1$ and $\hat{\theta}_2$ can be written as linear combinations of the coefficient matrices, $a(1)$ and $a(2)$ for, respectively, the two-treatment sub-designs, $D1$ and $D2$ defined in Section 6.1. As stated above we have

$$\begin{pmatrix}\hat{\theta}_1\\ \hat{\theta}_2\end{pmatrix} = \tilde{M}^{-1}\begin{bmatrix}a(1)\\ a(2)\end{bmatrix}, \tag{A4}$$

where the [] indicate that this should be interpreted as the linear combination of the matrices, i.e., that the coefficient matrix for $\hat{\theta}_1$ is $\tilde{m}^{11}a(1) + \tilde{m}^{12}a(2)$, where $\tilde{m}^{ij}$ are the elements of $\tilde{M}^{-1}$. The matrix $\tilde{M}$ is defined as

$$\begin{pmatrix}1 & \tilde{Q}_1\\ \tilde{Q}_2 & 1\end{pmatrix}\text{ (vertical analysis)}\quad\text{or}\quad\begin{pmatrix}1 & \frac{\tilde{Q}_1Q_{12}}{Q_1}\\ \frac{\tilde{Q}_2Q_{12}}{Q_2} & 1\end{pmatrix}\text{ (row-column).}$$

Here Q_1 is the value of Q the sub-design D1 and Q_2 is the corresponding quantity for D2. For three-treatment designs we also need Q_{12}, defined as

$$d_C^{12} = STG_{12} - TC_1^TC_2;\ \ d_{RC}^{12} = STG_{12} - SR_1^TR_2 - TC_1^TC_2 + G_1G_2;\ \text{ and }\ Q_{12} = d_{RC}^{12}/d_C^{12}.$$

Also with $d_C^1(d_C^2)$ defined as d_C for sub-design D1(D2), write $\tilde{Q}_1 = d_C^{12}/d_C^1$ and $\tilde{Q}_2 = d_C^{12}/d_C^2$.

Appendix 2: row-column designs with correlated rows

Two-treatment case

If $Y_s = (y_{s1}, y_{s2}, \ldots, y_{sT})^T$ is written for the mean of the N clusters on sequence s then $\mathsf{cov}(Y_s) = V/N$. If a model with column effects is fitted using (4), where $X_s = (I_T \mid D_{1s})$, with D_{1s} being the T-dimensional part of D_1 describing allocations in sequence s, then, using results for partitioned matrices once more we find

$$\hat{\theta} = \frac{S\sum_{s=1}^S D_{1s}^TV^{-1}Y_s - C_1^TV^{-1}Y_C}{S\sum_{s=1}^S D_{1s}^TV^{-1}D_{1s} - C_1^TV^{-1}C_1} \tag{A5}$$

with Y_C denoting the vector of $\sum_{s=1}^S y_{st}$, $t = 1, \ldots, T$.

If $V = \sigma^2((1-\psi)I_c + \psi J_{c,c})$ a formula for $\hat{\theta}$ can be derived by evaluating V^{-1} and substituting in (A5). Note that

$$V^{-1} = \frac{1}{\sigma^2(1-\psi)}[I_c - \tfrac{b}{T}J_{c,c}] \text{ with } b = \frac{T\psi}{1+\psi(T-1)}$$

In substituting this expression in (A5) the factor $\sigma^{-2}(1-\psi)^{-1}$ cancels, so can be ignored in the calculations. This gives

$$\begin{aligned}\hat{\theta} =& \frac{STy_{(1)} - TC_1^T Y_C - b[SR_1^T Y_R - G_1 Y_G]}{STG_1 - TC_1^T C_1 - b[SR_1^T R_1 - G_1^2]} \\ =& \frac{(1-b)[STy_{(1)} - TC_1^T Y_C] + b[STy_{(1)} - TC_1^T Y_C - SR_1^T Y_R + G_1 Y_G]}{(1-b)[STG_1 - TC_1^T C_1) + b[STG_1 - TC_1^T C_1 - SR_1^T R_1 + G_1^2]}.\end{aligned}$$

Comparing the elements of this expression with those for the vertical and row-column analyses, namely (A1) and (A2), it is readily seen that the above estimator can be written as a weighted mean of the vertical and row-column estimators, and that the coefficient matrix can be written as (5).

Three-treatment case

The result for three treatments proceeds analogously, giving

$$\begin{pmatrix}\hat{\theta}_1 \\ \hat{\theta}_2\end{pmatrix} = M^{-1}\begin{pmatrix}STy_{(1)} - TC_1^C y_C - b[SR_1^T y_S - G_1 y_G] \\ STy_{(2)} - TC_2^C y_C - b[SR_2^T y_S - G_2 y_G]\end{pmatrix} \tag{A6}$$

where

$$M = ST\begin{pmatrix}G_1 & G_{12} \\ G_{12} & G_2\end{pmatrix} - T\begin{pmatrix}C_1^T C_1 & C_1^T C_2 \\ C_1^T C_2 & C_2^T C_2\end{pmatrix} - b\left[S\begin{pmatrix}R_1^T R_1 & R_1^T R_2 \\ R_1^T R_2 & R_2^T R_2\end{pmatrix} - \begin{pmatrix}G_1^2 & G_1 G_2 \\ G_1 G_2 & G_2^2\end{pmatrix}\right]$$

If M_{ij} denotes the (i,j)th element of M, then the elements of the vector on the right hand side of (A6), divided, respectively, by M_{11} and M_{22}, are the estimators of θ_1, respectively θ_2, from the sub-designs $D1$ and $D2$. These can be re-cast as coefficient matrices and so provide the expression in (12).

Acknowledgement

We are grateful to Rosemary Bailey, Rob Henderson, Jim Hughes and the referees for their helpful comments on earlier versions of this paper.

References

1. Mdege ND, Man M, Taylor CA, Torgerson DJ. Systematic review of stepped wedge cluster randomized trials shows that design is particularly used to evaluate interventions during routine implementation. *Journal of Clinical Epidemiology* 2011; **64**:936–948, doi:10.1016/j.jclinepi.2010.12.003.
2. Beard E, Lewis JJ, Copas A, Davey C, Osrin D, Baio G, Thompson JA, Fielding KL, Omar RZ, Ononge S, *et al.*. Stepped wedge randomised controlled trials: systematic review of studies published between 2010 and 2014. *Trials* 2015; **16**, doi:10.1186/s13063-015-0839-2.
3. Girling AJ, Hemming K. Statistical efficiency and optimal design for stepped cluster studies under linear mixed effects models. *Statistics in Medicine* 2016; **35**:2149–2166, doi:10.1002/sim.6850.
4. Jones B, Kenward MG. *Design and Analysis of Cross-Over Trials*. 3rd edn., Chapman and Hall/CRCPress: Boca-Raton, 2015.
5. Davey C, Hargreaves J, Thompson JA, Copas AJ, Beard E, Lewis JJ, Fielding KL. Analysis and reporting of stepped wedge randomised controlled trials: synthesis and critical appraisal of published studies, 2010 to 2014. *Trials* 2015; **16**:358, doi:10.1186/s13063-015-0838-3.

6. Senn SJ. *Cross-Over Trials in Clinical Research*. 2nd edn., Wiley: Chichester, 2002, doi:10.1002/0470854596.
7. Copas AJ, Lewis JJ, Thompson JA, Davey C, Baio G, Hargreaves JR. Designing a stepped wedge trial: three main designs, carry-over effects and randomisation approaches. *Trials* 2015; **16**, doi:10.1186/s13063-015-0842-7.
8. Shah KR, Sinha BK. *The Theory of Optimal Designs*. Springer: Berlin, 1989.
9. Turner RM, White IR, Croudace T. Analysis of cluster randomized cross-over trial data: A comparison of methods. *Statistics in Medicine* 2007; **26**(2):274–289, doi:10.1002/sim.2537.
10. Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. *Contemporary Clinical Trials* 2007; **28**:182–191, doi:10.1016/j.cct.2006.05.007.
11. Hooper R, Teerenstra S, de Hoop E, Eldridge S. Sample size calculation for stepped wedge and other longitudinal cluster randomised trials. *Statistics in Medicine* 2016; **35**:4718–4728, doi:10.1002/sim.7028.
12. Lyons VH, Li L, Hughes JP, Rowhani-Rahbar A. Proposed variations of the stepped wedge design can be employed to accommodate multiple interventions. *Journal of Clinical Epidemiology* 2017; **??**:??, doi:10.1016/j.jclinepi.2017.04.004.
13. Reuther S, Holle D, Buscher I, Dortmann O, Mueller R, Bartholomeyczik S, Halek M. Effect evaluation of two types of dementia-specific case conferences in German nursing homes (FallDem) using a stepped-wedge design: study protocol for a randomized controlled trial. *Trials* 2014; **15**, doi:10.1186/1745-6215-15-319.
14. Patel MP, Sandige HL, Ndekha MJ, Briend A, Ashorn P, Manary MJ. Supplemental feeding with ready-to-use therapeutic food in malawian children at risk of malnutrition. *Journal of Health Population and Nutrition* 2005; **23**:351–357.
15. Haines TP, O'Brien L, Mitchell D, Bowles KA, Haas R, Markham D, Plumb S, Chiu T, May K, Philip K, *et al.*. Study protocol for two randomized controlled trials examining the effectiveness and safety of current weekend allied health services and a new stakeholder-driven model for acute medical/surgical patients versus no weekend allied health services. *Trials* 2015; **16**, doi:10.1186/s13063-015-0619-z.
16. Chinbuah MA, Kager PA, Abbey M, Gyapong M, Awini E, Nonvignon J, Adjuik M, Aikins M, Pagnoni F, Gyapong JO. Impact of community management of fever (using antimalarials with or without antibiotics) on childhood mortality: A cluster-randomized controlled trial in Ghana. *American Journal of Tropical Medicine and Hygiene* 2012; **87**:11–20, doi:10.4269/ajtmh.2012.12-0078.
17. Hughes JP, Granston TS, Heagerty PJ. Current issues in the design and analysis of stepped wedge trials. *Contemporary Clinical Trials* 2015; **45**:55–60, doi:10.1016/j.cct.2015.07.006.
18. Kershner RP, Federer WT. Two-treatment crossover designs for estimating a variety of effects. *Journal of the American Statistical Association* 1981; **76**:612–619, doi:10.2307/2287519.
19. Hemming K, Lilford R, Girling AJ. Stepped-wedge cluster randomised controlled trials: a generic framework including parallel and multiple-level designs. *Statistics in Medicine* 2015; **34**:181–196, doi:10.1002/sim.6325.
20. Matthews JNS. An optimal multi-period crossover design for an application in paediatric nephrology. *Statistics in Medicine* 2014; **33**:857–866, doi:10.1002/sim.5981.
21. Matthews JNS. Recent developments in crossover designs. *International Statistical Review* 1988; **56**:117–127, doi:10.2307/1403636.
22. Fleiss JL. On multiperiod crossover studies. *Biometrics* 1986; **42**:449–450.
23. Rhoda DA, Murray DM, Andridge RR, Pennell ML, Hade EM. Studies with staggered starts: multiple baseline designs and group-randomized trials. *American Journal of Public Health* 2011; **101**:2164–2169.